

IS THE MULTIGRID METHOD FAULT TOLERANT? THE MULTILEVEL CASE.

MARK AINSWORTH AND CHRISTIAN GLUSA

ABSTRACT. Computing at the exascale level is expected to be affected by a significantly higher rate of faults, due to increased component counts as well as power considerations. Therefore, current day numerical algorithms need to be reexamined as to determine if they are fault resilient, and which critical operations need to be safeguarded in order to obtain performance that is close to the ideal fault-free method.

In a previous paper [1], a framework for the analysis of random stationary linear iterations was presented and applied to the two grid method. The present work is concerned with the multigrid algorithm for the solution of linear systems of equations, which is widely used on high performance computing systems. It is shown that the Fault-Prone Multigrid Method is not resilient, unless the prolongation operation is protected. Strategies for fault detection and mitigation as well as protection of the prolongation operation are presented and tested, and a guideline for an optimal choice of parameters is devised.

1. INTRODUCTION

The anticipated arrival of exascale systems opens up the possibility for scientific simulation at dramatically increased scale, provided that several technical challenges can be overcome. One such challenge concerns the resilience of the underlying numerical algorithm in the face of the increased level of faults and node failures on an exascale machine [6, 7].

Faults are classified as *hard* or *soft* [2]. While hard faults require immediate remedial action in order for the program execution to proceed, the effect of soft faults is to perturb data and instructions, possibly undetected but having the potential to degrade the performance of the algorithm or even invalidate the entire simulation. Node failures or lost messages constitute examples of hard faults, whereas random bit flips, induced by cosmic particles, can be classified as soft faults. It is of urgent concern to determine the effect of faults on existing state of the art numerical methods which will be utilised on exascale systems. In the cases where the algorithms are found wanting, one would like to know how to modify the schemes to cope with the new reality of fault-prone computing.

The widely used *multigrid algorithm* for the solution of linear systems of equations is the concern of this work. In earlier work [1], we introduced a simple fault mitigation strategy called *laissez-faire*, whereby, instead of trying to recover any lost or corrupted information, one simply replaces affected values with zero before continuing the execution. This approach has several advantages in terms of

1991 *Mathematics Subject Classification.* 65F10, 65N22, 65N55, 68M15.

Key words and phrases. Multigrid, Fault Tolerance, Resilience, Random Matrices, Convergence Analysis.

The effort of MA was partially supported by SIRIUS award from DoE..

efficiency including no requirement for communication. However, the acid test is whether or not this strategy is effective and results in any improvement in fault resilience.

The goal of this work is twofold. Firstly, we extend the analytic convergence estimates from the two grid setting to an arbitrary number of levels. Secondly, we investigate how the algorithm can be applied in practice where some form of fault detection is needed. In Section 2, we will recall the basic theoretical framework for fault-prone linear iterative methods as well as some previous results needed in the analysis. We then state a general convergence estimate for Fault-Prone Multigrid in Section 3, and demonstrate its validity for a range of numerical examples. Finally, the practical issue of fault detection and protection of the involved operations is discussed in Section 4. The effects of different levels of protection are demonstrated for a model problem, and optimal parameters depending on fault rate and problem size are given. The theoretical details and proofs of the results are collected in the Appendix.

2. PRELIMINARIES

2.1. Modelling Faults. In [1], we proposed to model the effect of faults and their mitigation in iterative linear solvers through random diagonal matrices. A fault-free vector $x \in \mathbb{R}^n$ gets transformed into \tilde{x} according to

$$\tilde{x} = \mathcal{X}x.$$

Throughout an algorithm, several operations might be subject to faults, each corresponding to such a fault matrix. For given $\varepsilon > 0$, let \mathbb{S}_ε denote the set of these fault matrices, and assume that they satisfy the following conditions:

Assumption 1.

- (1) Each $\mathcal{X} \in \mathbb{S}_\varepsilon$ is a random diagonal matrix.
- (2) For every $\mathcal{X} \in \mathbb{S}_\varepsilon$, there holds $\mathbb{E}[\mathcal{X}] = e(\mathcal{X})\mathbf{I}$, where $e(\mathcal{X}) > 0$, and $|e(\mathcal{X}) - 1| \leq C\varepsilon$ for some fixed $C > 0$.
- (3) For every $\mathcal{X} \in \mathbb{S}_\varepsilon$ there holds $\|\mathbb{V}[\mathcal{X}]\|_2 = \max_{i,j} |\text{Cov}(\mathcal{X}_{ii}, \mathcal{X}_{jj})| \leq \varepsilon$.

ε measures how close the fault matrices are to the identity, i.e. the fault-free case. While the value of ε could vary from operation to operation, for example to take into account that denser matrix-vector products take more time and are therefore more likely to be hit by faults, we neglect this aspect in the analysis for the sake of clarity.

Important examples of random faults covered by Assumption 1 are

- (1) *Componentwise detectable faults*

A typical example of a detectable fault would be flipping of individual bits in a floating point number, resulting in a large enough upset to be detected, or a pointer corruption that leads to an invalid memory address. Such cases can be treated using the laissez-faire strategy described above, and therefore modelled as componentwise faults with

$$(1) \quad \mathcal{X} = \text{diag}(\chi_1, \dots, \chi_n),$$

where χ_i are independent and identically ε -distributed Bernoulli random variables, i. e.

$$\chi_i = \begin{cases} 0 & \text{with probability } \varepsilon, \\ 1 & \text{with probability } 1 - \varepsilon. \end{cases}$$

(2) *Blockwise detectable faults*

In the event of a node failure and application of the laissez-faire strategy, all the components of a vector that were residing on the node will be zeroed out. This can be modelled by a random matrix with independent diagonal blocks:

$$(2) \quad \mathcal{X} = \text{blockdiag}(\chi_1 \mathbf{I}_1, \dots, \chi_N \mathbf{I}_N),$$

where χ_i are independent and identically ε -distributed Bernoulli random variables and \mathbf{I}_j are identity matrices.

(3) *Silent faults*

Soft faults may be difficult or even impossible to detect, especially if their induced perturbation is small. Such silent faults can be modelled by a random matrix

$$(3) \quad \mathcal{X} = \mathbf{I} + \text{diag}(\eta_1 \chi_1, \dots, \eta_n \chi_n)$$

where η_i are independent and identically distributed random variables, and χ_i are independent and identically distributed Bernoulli random variables, such that $\mathcal{X} \in \mathbb{S}_\varepsilon$. In particular, this includes the cases of frequent faults with small impact and of rare but large upsets.

2.2. Multigrid Algorithm. Let \mathbf{A} be a symmetric, positive definite matrix arising from a finite element discretization of an elliptic partial differential equation in d spatial dimensions. The multigrid method solves the linear system of equations

$$\mathbf{A}x = b$$

by introducing a hierarchy of coarser problems

$$\mathbf{A}_\ell x_\ell = b_\ell$$

of size n_ℓ , $\ell = 0, \dots, L$ with $\mathbf{A}_L = \mathbf{A}$. Information gets transferred between levels through restriction and prolongation operators

$$\mathbf{R}_\ell^{\ell+1} : \mathbb{R}^{n_{\ell+1}} \rightarrow \mathbb{R}^{n_\ell}, \quad \mathbf{P}_{\ell+1}^\ell : \mathbb{R}^{n_\ell} \rightarrow \mathbb{R}^{n_{\ell+1}}.$$

We will assume that $\mathbf{P}_{\ell+1}^\ell = \left(\mathbf{R}_\ell^{\ell+1}\right)^T$ along with the usual Galerkin relation $\mathbf{A}_\ell = \mathbf{R}_\ell^{\ell+1} \mathbf{A}_{\ell+1} \mathbf{P}_{\ell+1}^\ell$. We will drop sub- and superscripts on restriction and prolongation operators in what follows. Moreover, smootheners are defined on each level as

$$x_\ell \leftarrow x_\ell + \mathbf{N}_\ell (b_\ell - \mathbf{A}_\ell x_\ell),$$

where \mathbf{N}_ℓ is a suitable preconditioner, e.g. the damped Jacobi preconditioner $\mathbf{N}_\ell = \theta \mathbf{D}_\ell^{-1}$, with \mathbf{D}_ℓ the diagonal part of \mathbf{A}_ℓ .

The Fault-Prone Multigrid Method \mathcal{MG}_ℓ was described in detail in [1], and is given by Algorithm 1. The classical fault-free variant MG_ℓ can be obtained by replacing all fault matrices $\mathcal{X}_\ell^{(\bullet)}$ with identity matrices.

Algorithm 1 Model for Fault-Prone Multigrid Algorithm \mathcal{MG}_ℓ where $\mathcal{X}_\ell^{(\bullet)}$ are random diagonal matrices.

Input: Right hand side b_ℓ ; Initial iterate x_ℓ

Output: $\mathcal{MG}_\ell(b_\ell, x_\ell)$

```

1: if  $\ell = 0$  then return  $\mathbf{A}_0^{-1}b_0$                                  $\triangleright$  Exact solve on coarsest grid
2: else
3:   for  $i \leftarrow 1$  to  $\nu_1$  do
4:      $x_\ell \leftarrow x_\ell + \mathcal{X}_\ell^{(S, \text{pre})} \mathbf{N}_\ell (b_\ell - \mathbf{A}_\ell x_\ell)$            $\triangleright \nu_1$  pre-smoothing steps
5:      $d_{\ell-1} \leftarrow \mathcal{X}_{\ell-1}^{(R)} \mathbf{R} \mathcal{X}_\ell^{(\rho)} (b_\ell - \mathbf{A}_\ell x_\ell)$            $\triangleright$  Restriction to coarser grid
6:      $e_{\ell-1} \leftarrow 0$ 
7:     for  $j \leftarrow 1$  to  $\gamma$  do
8:        $e_{\ell-1} \leftarrow \mathcal{MG}_{\ell-1}(d_{\ell-1}, e_{\ell-1})$            $\triangleright \gamma$  coarse grid correction steps
9:        $x_\ell \leftarrow x_\ell + \mathcal{X}_\ell^{(P)} \mathbf{P} e_{\ell-1}$            $\triangleright$  Prolongation to finer grid
10:    for  $i \leftarrow 1$  to  $\nu_2$  do
11:       $x_\ell \leftarrow x_\ell + \mathcal{X}_\ell^{(S, \text{post})} \mathbf{N}_\ell (b_\ell - \mathbf{A}_\ell x_\ell)$            $\triangleright \nu_2$  post-smoothing steps
```

The classical approach to the analysis of iterative solution methods for linear systems uses the notion of an iteration matrix. For the fault-prone method \mathcal{MG}_ℓ , it is defined by the equation

$$x - \mathcal{MG}_\ell(\mathbf{A}_\ell x, y) = \mathcal{E}_\ell(x - y), \quad \forall x, y \in \mathbb{R}^{n_\ell},$$

and given by

$$\mathcal{E}_0 = \mathbf{0},$$

$$(4) \quad \mathcal{E}_\ell = \left(\mathcal{E}_\ell^{S, \text{post}} \right)^{\nu_2} \left[\mathbf{I} - \mathcal{X}_\ell^{(P)} \mathbf{P} (\mathbf{I} - \mathcal{E}_{\ell-1}^\gamma) \mathbf{A}_{\ell-1}^{-1} \mathcal{X}_{\ell-1}^{(R)} \mathbf{R} \mathcal{X}_\ell^{(\rho)} \mathbf{A}_\ell \right] \left(\mathcal{E}_\ell^{S, \text{pre}} \right)^{\nu_1},$$

for $\ell = 1, \dots, L$, with $\mathcal{E}_\ell^{S, \text{pre/post}} = \mathbf{I} - \mathcal{X}_\ell^{(S, \text{pre/post})} \mathbf{N}_\ell \mathbf{A}_\ell$. Here, we have used superscripts pre and post to reflect that the pre- and post-smootheners are *independent realisations of the same random matrix*. Moreover, we will use powers of random matrices to signify products of independent realisations of the same random matrix.

Setting $\mathcal{E}_{L-1} = \mathbf{0}$ and applying the recursion (4) in the case $\ell = L$ yields a formula for the iteration matrix of the *Fault-Prone Two Grid Algorithm*:

$$(5) \quad \mathcal{E}_L^{TG} = \left(\mathcal{E}_L^{S, \text{post}} \right)^{\nu_2} \left[\mathbf{I} - \mathcal{X}_L^{(P)} \mathbf{P} \mathbf{A}_{L-1}^{-1} \mathcal{X}_{L-1}^{(R)} \mathbf{R} \mathcal{X}_L^{(\rho)} \mathbf{A}_L \right] \left(\mathcal{E}_L^{S, \text{pre}} \right)^{\nu_1}$$

$$= \left(\mathcal{E}_L^{S, \text{post}} \right)^{\nu_2} \mathcal{E}_L^{CG} \left(\mathcal{E}_L^{S, \text{pre}} \right)^{\nu_1},$$

corresponding to using an exact solver on level $L - 1$. Here \mathcal{E}_ℓ^{CG} is the iteration matrix of the exact fault-prone coarse grid correction on level ℓ .

By replacing all fault matrices with the identity, the classical fault-free quantities are recovered:

$$\mathbf{E}_\ell = \left(\mathbf{E}_\ell^S \right)^{\nu_2} \left[\mathbf{I} - \mathbf{P} (\mathbf{I} - \mathbf{E}_{\ell-1}^\gamma) \mathbf{A}_{\ell-1}^{-1} \mathbf{R} \mathbf{A}_\ell \right] \left(\mathbf{E}_\ell^S \right)^{\nu_1},$$

$$\mathbf{E}_\ell^S = \mathbf{I} - \mathbf{N}_\ell \mathbf{A}_\ell,$$

$$\mathbf{E}_\ell^{TG} = \left(\mathbf{E}_\ell^S \right)^{\nu_2} \mathbf{E}_\ell^{CG} \left(\mathbf{E}_\ell^S \right)^{\nu_1},$$

$$\mathbf{E}_\ell^{CG} = \mathbf{I} - \mathbf{P}\mathbf{A}_\ell^{-1}\mathbf{R}\mathbf{A}_\ell.$$

These are the iteration matrices of the fault-free multigrid methods, smootheners, two grid method and coarse-grid correction.

2.3. Convergence Theory for Standard Multigrid. The convergence proof of the Fault-Prone Multigrid Method is motivated by the classical analysis of the fault-free algorithm. The standard assumptions for the convergence analysis of the fault-free multigrid method read as follows [5, 11, 12, 17]:

Assumption 2 (Smoothing property). *There exists $\eta : \mathbb{N} \rightarrow \mathbb{R}_{\geq 0}$ satisfying $\lim_{\nu \rightarrow \infty} \eta(\nu) = 0$ and such that*

$$\left\| \mathbf{A}_\ell \left(\mathbf{E}_\ell^S \right)^\nu \right\|_2 \leq \eta(\nu) \|\mathbf{A}_\ell\|_2, \quad \nu \geq 0, \ell = 1, \dots, L.$$

Assumption 3 (Approximation property). *There exists a constant C_A such that*

$$\left\| \mathbf{E}_\ell^{CG} \mathbf{A}_\ell^{-1} \right\|_2 \leq \frac{C_A}{\|\mathbf{A}_\ell\|_2}, \quad \ell = 1, \dots, L.$$

The smoothing and approximation property imply two grid convergence, with convergence rate given by

$$\begin{aligned} (6) \quad \rho \left(\mathbf{E}_\ell^{TG} \right) &\leq \left\| \mathbf{E}_\ell^{CG} \left(\mathbf{E}_\ell^S \right)^{\nu_1 + \nu_2} \right\|_2 \\ &\leq \left\| \mathbf{E}_\ell^{CG} \mathbf{A}_\ell^{-1} \right\|_2 \left\| \mathbf{A}_\ell \left(\mathbf{E}_\ell^S \right)^{\nu_1 + \nu_2} \right\|_2 \\ &\leq C_A \eta(\nu_1 + \nu_2). \end{aligned}$$

The right-hand side is less than one provided $\nu_1 + \nu_2$ is large enough.

Assumption 4. *The smootheners are non-expansive, i.e. $\rho \left(\mathbf{E}_\ell^S \right) = \left\| \mathbf{E}_\ell^S \right\|_A \leq 1$, and there exists $C_S > 0$ such that*

$$\left\| \left(\mathbf{E}_\ell^S \right)^\nu \right\|_2 \leq C_S \quad \nu \geq 1, \ell = 1, \dots, L.$$

Assumption 4 permits to show that the two grid method is also a contraction with respect to the spectral norm $\|\bullet\|_2$:

$$\left\| \mathbf{E}_\ell^{TG} \right\|_2 \leq \left\| \left(\mathbf{E}_\ell^S \right)^{\nu_2} \right\|_2 \left\| \mathbf{E}_\ell^{CG} \left(\mathbf{E}_\ell^S \right)^{\nu_1} \right\|_2 \leq C_S C_A \eta(\nu_1).$$

While this result is weaker than (6), it will be useful in the fault-prone case.

Assumption 5. *There exist positive constants \underline{C}_p and \overline{C}_p such that*

$$\underline{C}_p^{-1} \|x_\ell\|_2 \leq \|\mathbf{P}x_\ell\|_2 \leq \overline{C}_p \|x_\ell\|_2 \quad \forall x_\ell \in \mathbb{R}^{n_\ell}, \ell = 0, \dots, L-1.$$

Assumptions 4 and 5 allow to extend the convergence theory to the multilevel case with $\gamma \geq 2$. The most interesting case in practice is the W-cycle ($\gamma = 2$).

$$\begin{aligned} \|\mathbf{E}_\ell\|_2 &= \left\| \left(\mathbf{E}_\ell^S \right)^{\nu_2} \left[\mathbf{I} - \mathbf{P} \left(\mathbf{I} - \mathbf{E}_{\ell-1}^\gamma \right) \mathbf{A}_{\ell-1}^{-1} \mathbf{R} \mathbf{A}_\ell \right] \left(\mathbf{E}_\ell^S \right)^{\nu_1} \right\|_2 \\ &\leq \left\| \left(\mathbf{E}_\ell^S \right)^{\nu_2} \left[\mathbf{I} - \mathbf{P} \mathbf{A}_{\ell-1}^{-1} \mathbf{R} \mathbf{A}_\ell \right] \left(\mathbf{E}_\ell^S \right)^{\nu_1} \right\|_2 \\ &\quad + \left\| \left(\mathbf{E}_\ell^S \right)^{\nu_2} \right\|_2 \|\mathbf{P}\|_2 \|\mathbf{E}_{\ell-1}\|_2^\gamma \left\| \mathbf{A}_{\ell-1}^{-1} \mathbf{R} \mathbf{A}_\ell \left(\mathbf{E}_\ell^S \right)^{\nu_1} \right\|_2 \end{aligned}$$

$$\leq \left\| \mathbf{E}_\ell^{TG} \right\|_2 + \underline{C}_p \overline{C}_p C_S \left\| \mathbf{E}_{\ell-1} \right\|_2^\gamma \left\| \mathbf{P} \mathbf{A}_{\ell-1}^{-1} \mathbf{R} \mathbf{A}_\ell \left(\mathbf{E}_\ell^S \right)^{\nu_1} \right\|_2.$$

Since

$$\begin{aligned} \left\| \mathbf{P} \mathbf{A}_{\ell-1}^{-1} \mathbf{R} \mathbf{A}_\ell \left(\mathbf{E}_\ell^S \right)^{\nu_1} \right\|_2 &\leq \left\| \left(\mathbf{I} - \mathbf{P} \mathbf{A}_{\ell-1}^{-1} \mathbf{R} \mathbf{A}_\ell \right) \left(\mathbf{E}_\ell^S \right)^{\nu_1} \right\|_2 + \left\| \left(\mathbf{E}_\ell^S \right)^{\nu_1} \right\|_2 \\ &\leq C_S + C_A \eta(\nu_1), \end{aligned}$$

we obtain the recursive inequality

$$(7) \quad \left\| \mathbf{E}_\ell \right\|_2 \leq \left\| \mathbf{E}_\ell^{TG} \right\|_2 + C_* \left\| \mathbf{E}_{\ell-1} \right\|_2^\gamma$$

with $C_* = C_S \underline{C}_p \overline{C}_p (C_S + C_A \eta(\nu_1))$. A classical result [11] concerning this inequality is

Lemma 1. *Suppose the elements of the sequence $\{\eta_\ell\}_{\ell \geq 1}$ satisfy $0 \leq \eta_1 \leq \xi$ and $0 \leq \eta_\ell \leq \xi + C_* \eta_{\ell-1}^\gamma$, $\ell \geq 2$. If $\gamma \geq 2$, $C_* \gamma > 1$ and $\xi \leq \frac{\gamma-1}{\gamma} \frac{1}{(C_* \gamma)^{\frac{1}{\gamma-1}}}$, then*

$$\eta_\ell \leq \begin{cases} \frac{\gamma}{\gamma-1} \xi, & \gamma \geq 2, \\ \frac{2\xi}{1+\sqrt{1-4C_* \xi}}, & \gamma = 2 \end{cases} < 1.$$

The result show that two grid convergence along with a sufficient number of smoothing steps imply the multilevel scheme is convergent in the absence of faults.

2.4. Review of Previous Work on the Fault-Prone Two Grid Method. The iteration matrix of the Fault-Prone Multigrid Method is random, and the usual the spectral radius used in the fault-free case is no longer relevant. Instead, it transpires that the asymptotic rate of convergence in the fault-prone case is governed by the *Lyapunov spectral radius* of the iteration matrix:

$$\varrho(\mathcal{E}_L) = \lim_{N \rightarrow \infty} \exp \left\{ \mathbb{E} \left[\log \left\| \mathcal{E}_L^N \right\|^{1/N} \right] \right\}.$$

We refer the reader to [1, 4, 8] for further discussion and details relating to the Lyapunov spectral radius. In particular, [1] describes the so called *Replica trick* which gives the following bound for the Lyapunov spectral radius

$$(8) \quad \varrho(\mathcal{E}_L) \leq \sqrt{\rho(\mathbb{E}[\mathcal{E}_L \otimes \mathcal{E}_L])}.$$

Under suitable assumptions, it was shown [1] that the Lyapunov spectral radius for the Fault-Prone Two Grid Method satisfies

$$(9) \quad \varrho(\mathcal{E}_L^{TG}) \leq \left\| \mathbf{E}_L^{TG} \right\|_A + C \begin{cases} \varepsilon n_L^{(4-d)/2d} & d < 4, \\ \varepsilon \sqrt{\log n_L} & d = 4, \\ \varepsilon & d > 4, \end{cases}$$

where d is the spatial dimension of the underlying PDE, and $\|\bullet\|_A$ is the usual energy norm. The estimate suggests, as confirmed by numerical examples, that the convergence rate of the Fault-Prone Two Grid Method degenerates as $n_L \rightarrow \infty$ and eventually fails to converge. Moreover, it can be shown [1] that protection of the prolongation operation against faults (so that $\mathcal{X}_L^{(P)} = \mathbf{I}$), whilst allowing other sources of faults to remain, results in the Two Grid scheme being resilient:

Theorem 2. Let $\mathcal{E}_\ell^{TG}(\nu_1, \nu_2) = \left(\mathcal{E}_\ell^{S,post}\right)^{\nu_2} \mathcal{E}_\ell^{CG} \left(\mathcal{E}_\ell^{S,pre}\right)^{\nu_1}$ be the iteration matrix of the Fault-Prone Two Grid Method with faults in smoother, residual and restriction, and protected prolongation. Provided Assumptions 1 to 5 hold, and that \mathbf{N}_ℓ and $\mathcal{X}_\ell^{(S)}$ commute, we find that

$$\varrho\left(\mathcal{E}_\ell^{TG}(\nu_1, \nu_2)\right) \leq \left\| \mathbb{E} \left[\left(\mathcal{E}_\ell^{TG}(\nu_1, \nu_2)\right)^{\otimes 2} \right] \right\|_{A^2}^{1/2} \leq \left\| \mathbf{E}_\ell^{TG}(\nu_2, \nu_1) \right\|_2 + C\varepsilon,$$

where the constant C is independent of ε and ℓ , and $\|\bullet\|_{A^2}$ is the double energy norm defined in the Appendix.

Theorem 2 shows that the convergence of the Fault-Prone Two Grid scheme with a protected prolongation does not degenerate as $n_L \rightarrow \infty$. One of the main aims of the present work is to generalise Theorem 2 to the case of the Fault-Prone Multigrid Method.

3. MAIN RESULTS

Consider a second order elliptic PDE with homogeneous Dirichlet boundary conditions on a polyhedral domain Ω given in variational form by

$$\text{Find } u \in V = H_0^1(\Omega) : a(u, v) = L(v) \quad \forall v \in V.$$

Here, the bilinear form a is continuous and V -coercive, and the linear form L is continuous. Using a shape regular partitioning \mathcal{T}_L of Ω , the finite dimensional space of continuous piecewise polynomials of order k is defined as

$$V_L = \{v \in H_0^1(\Omega) \mid v|_K \in \mathbb{P}_k(K) \quad \forall K \in \mathcal{T}_L\}.$$

Letting ϕ_L be the vector of nodal basis functions $\phi_L^{(i)}$, $i = 1, \dots, n_L$, the solution $u_L \in V_L$ to the PDE is approximated by

$$\mathbf{A}_L x_L = b_L,$$

where $\mathbf{A}_L = a(\phi_L, \phi_L)$, $b_L = L(\phi_L)$, and $u_L = \phi_L \cdot x_L$.

The hierarchy of levels for the multigrid solver is constructed from discretizations of the same problem on nested coarser meshes \mathcal{T}_ℓ . Uniform refinement [15] or adaptive mesh refinement can be used to obtain successively finer meshes, and the canonical injection of the coarser space into the finer one is used to define prolongation and restriction.

3.1. Convergence Estimate for the Fault-Prone W-Cycle. While the two grid method is, as a solver, mostly of academic interest, the behaviour of the multigrid method under the impact of faults is of great practical importance. We extend the result of Theorem 2 to the W-cycle. This theorem mirrors the classical implication of W-cycle convergence by two grid convergence, but applies to the Lyapunov spectral radius needed for the analysis of the Fault-Prone Multigrid Method. No additional assumptions are required beyond these needed for the classical multigrid analysis.

Theorem 3. Provided Assumptions 1 to 5 hold, that the prolongation is protected, that the number of smoothing steps is sufficient, and that ε is sufficiently small, the

fault-prone multigrid method converges with a rate bounded by

$$\varrho(\mathcal{E}_\ell(\nu_1, \nu_2, \gamma)) \leq \begin{cases} \frac{\gamma}{\gamma-1} C_{TG} + C\varepsilon, & \gamma \geq 2, \\ \frac{2}{1+\sqrt{1-4C_*C_{TG}}} C_{TG} + C\varepsilon, & \gamma = 2, \end{cases}$$

where C is independent of ε and ℓ and

$$C_{TG} = \max_{\ell} \left\| \mathbf{E}_\ell^{TG}(\nu_2, \nu_1) \right\|_2 \leq C_S C_A \eta(\nu_2),$$

$$C_* = C_S \underline{C}_p \overline{C}_p (C_S + C_A \eta(\nu_2)).$$

Setting the fault rate $\varepsilon = 0$ in the above bounds recovers the classical estimates from Section 2 for the fault-free multigrid method, since $\varrho(\mathcal{E}_\ell)$ reduces to $\rho(\mathbf{E}_\ell)$ for $\varepsilon = 0$. Just as the Two Grid result, this Theorem makes no assumptions about the origin of the solver hierarchy, and does not rely on a particular choice of smoothener. We refer the reader to Appendix A for the proof of Theorem 3.

3.2. The Effect of Protection of the Prolongation. Theorem 3 assumes that the prolongation operator is protected. We investigate whether this assumption can be relaxed by considering a numerical example where we do not protect the prolongation. Specifically, we consider the Poisson equation in two dimensions

$$(10) \quad \begin{cases} -\Delta u = f & \text{in } \Omega = [0, 1]^2, \\ u = 0 & \text{on } \partial\Omega, \end{cases}$$

and in order to rule out extraneous effects, we use a uniform mesh, a discretization with piecewise linear finite elements, and optimally damped Jacobi pre- and post-smootheners. In particular, it is well established that the fault-free multigrid method converges in this setting. Assumptions 2 to 5 are satisfied, as for example shown in [12]. On every level, residual, restriction, prolongation and smootheners are subject to componentwise faults, as given in (1). We consider problems of size between 1 million and 1 billion degrees of freedom, and fault probabilities between 10^{-4} and 0.6. To minimize floating point contamination in the approximation of the Lyapunov spectral radius, the right-hand side f is taken to be zero, a non-zero random initial iterate is chosen, and after each iteration the current iterate is renormalised. In Figure 1, we plot the evolution of the residual norm with respect to the iteration number for the case of laissez-faire mitigation in prolongation, restriction, residual and smootheners. We can see that the number of degrees of freedom n_L adversely affects the rate of convergence, even leading to divergence for large number of unknowns. Estimates of the Lyapunov spectral radius are obtained as the geometric average of 1000 iterations of Fault-Prone Multigrid, and are displayed in Figure 1.

In the Two Grid setting which was discussed in [1], we found that the asymptotic rate for this problem scales like $\sqrt{n_L \varepsilon^2}$. Similarly, as seen in Figure 1, we find

$$(11) \quad \varrho(\mathcal{E}_L) = C_0 + \mathcal{O}\left(\sqrt{n_L \varepsilon^2}\right)$$

for the multilevel case, with C_0 being a constant related to the fault-free method. This means that the method is not fault resilient, since for any given rate of faults ε , there exists a maximum problem size above which multigrid diverges. Therefore additional protection is mandatory in the multilevel case as well.

We repeat the same experiments, this time with a protected prolongation, i.e. $\mathcal{X}_\ell^{(P)} = \mathbf{I}$. This produces the desired independence of the problem size, as predicted

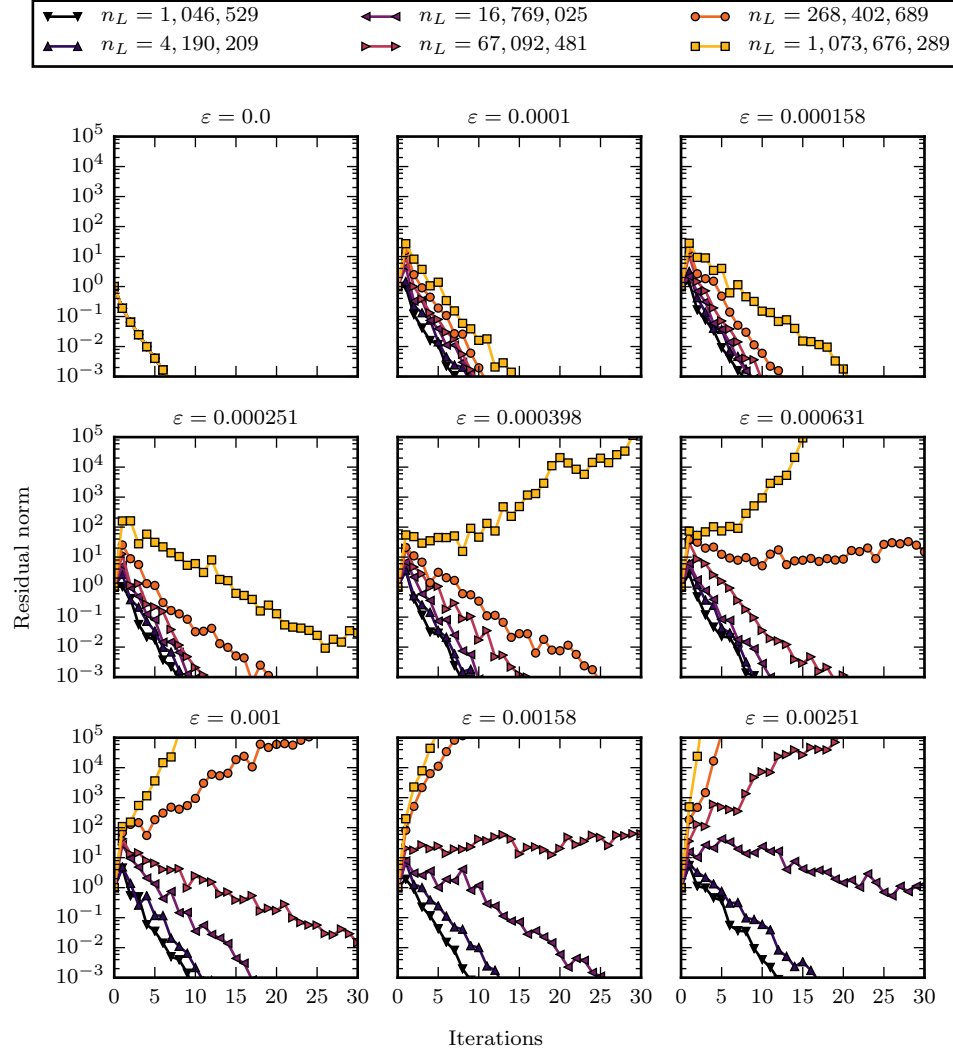


FIGURE 1. Plots of the norm of the residual against iteration number for the Fault-Prone W-Cycle Multigrid Method in the case of discretization of Poisson problem on square domain.

by Theorem 3 and as shown by the evolution of the residual in Figure 3 and by the estimated rate of convergence in Figure 1. We further illustrate the results with several test problems that pose more of a challenge to the multigrid solver.

3.3. Adaptively Refined Meshes. A second numerical example illustrates the results of Theorem 3 for the case of an adaptively refined mesh. We consider the 2D magnetostatics problem for a three phase 6/4 switched reluctance motor as

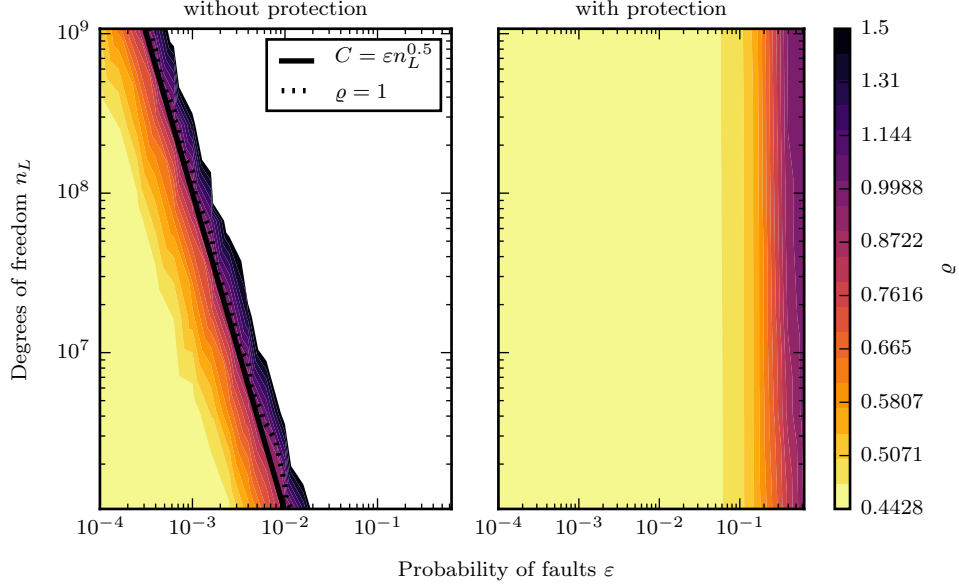


FIGURE 2. Lyapunov spectral radius $\rho(\mathcal{E}_L)$ for the iteration matrix for the Fault-Prone W-Cycle Multigrid method in the case of discretization of Poisson problem on a square domain. Without protected prolongation (*left*), and with protected prolongation (*right*).

depicted in Figure 4. Gauss's law and Ampère's law are given by

$$\operatorname{div} \vec{B} = 0, \quad \operatorname{curl} \vec{H} = J,$$

where \vec{B} is the magnetic flux density, \vec{H} the magnetic field intensity and J the current density. \vec{B} and \vec{H} are linked through the constitutive relation $\vec{B} = \mu \vec{H}$ with magnetic permeability μ . Using the magnetic vector potential $\vec{B} = \operatorname{curl} u$, the system can be rewritten as

$$-\frac{1}{\mu} \Delta u = J.$$

This gives rise to a variational problem with

$$a(u, v) = \int_{\Omega} \frac{1}{\mu} \operatorname{grad} u \cdot \operatorname{grad} v, \quad F(v) = \int_{\Omega} Jv.$$

The permeability μ is 5200 in the rotor and the stator, and 1 everywhere else. The current density J is unity in the coils and 0 everywhere else. Using continuous piecewise linear finite elements and classical residual-based local a posteriori error indicators [10], we adaptively refine an initial mesh shown in Figure 4. We use Red-Green refinement [3] coupled with a Dörfler marking strategy [9] with refinement parameter 0.5. The solver hierarchy is generated without injection of faults; then the problem is solved for componentwise faults and varying fault rates.

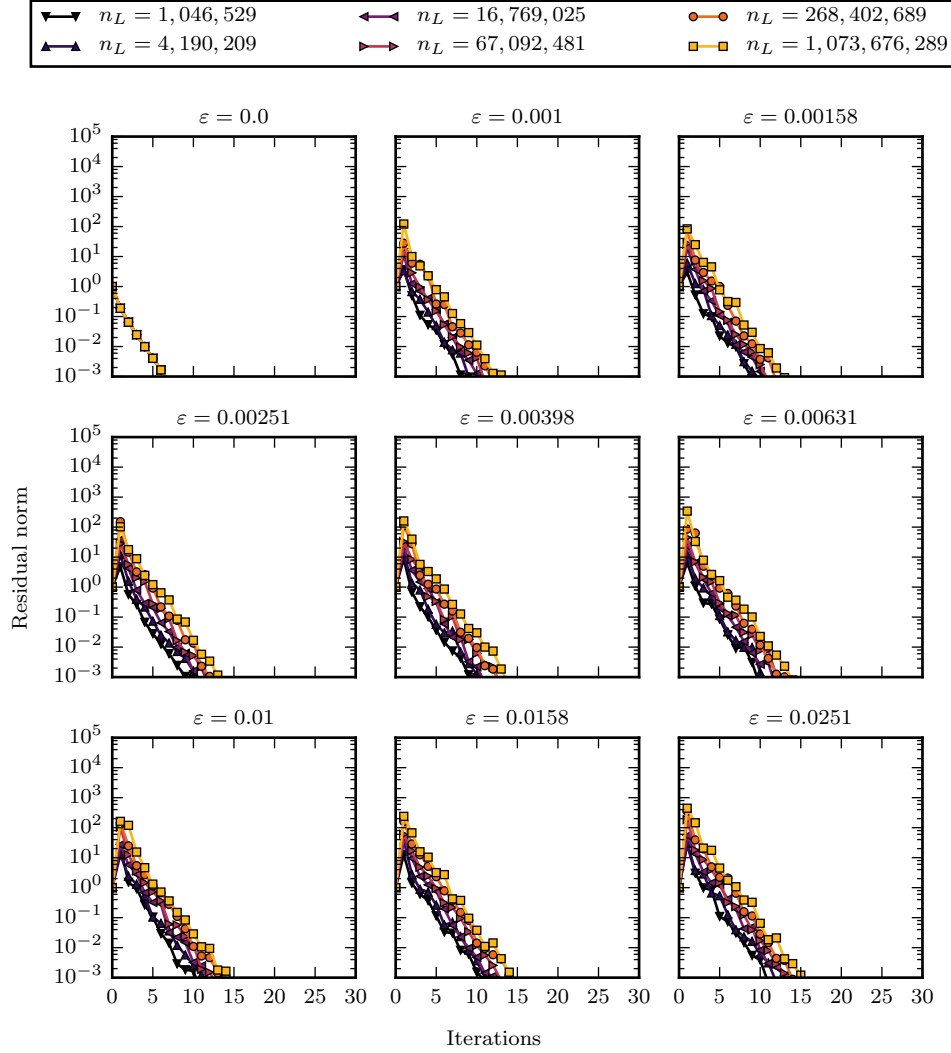


FIGURE 3. Plot of the norm of the residual against iteration number for the Fault-Prone W-Cycle Multigrid Method with protected prolongation in the case of discretization of Poisson problem on square domain.

Figure 4 depicts the approximate rate of convergence of the W-Cycle with one step of Jacobi pre- and post-smoothing each and without protection of the prolongation. We can see that although neither the W-Cycle nor the adaptively refined mesh are covered by the two grid result from [1], the convergence estimate is qualitatively correct. Added protection of the prolongation operation leads to a fault resilient method, as shown in Figure 4. Small variations with respect to the number

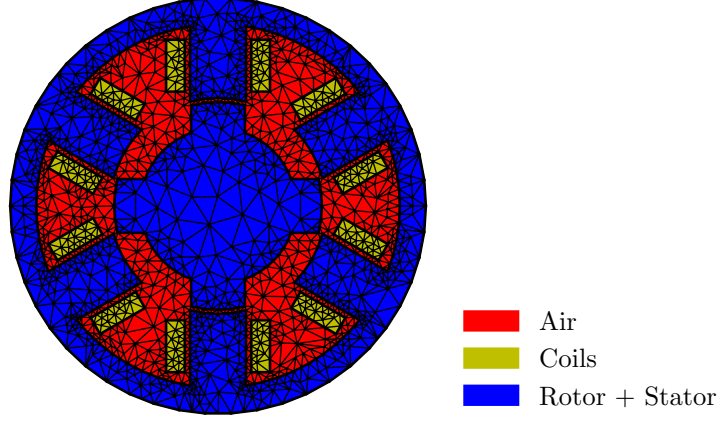
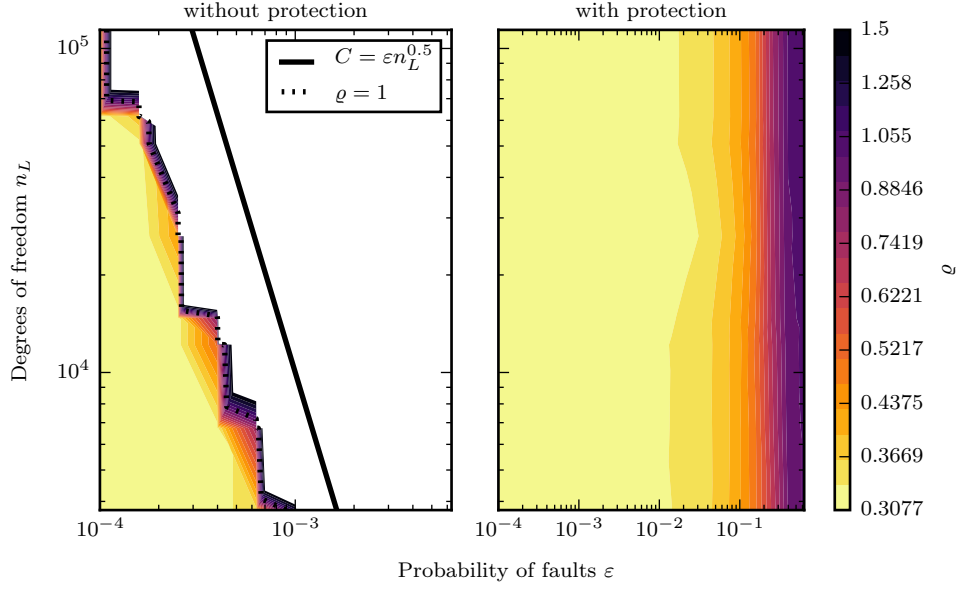


FIGURE 4. Initial mesh for the motor problem.

FIGURE 5. Lyapunov spectral radius $\rho(\mathcal{E}_L)$ for the iteration matrix for the Fault-Prone W-Cycle Multigrid Method in the case of discretization of the motor problem. Without protected prolongation (*left*), and with protected prolongation (*right*).

of degrees of freedom are due to varying coarsening ratios of the levels. The results match Theorem 3, even though Assumption 3 is not satisfied.

3.4. Higher Spatial Dimension and Higher Order Finite Elements. We now demonstrate that fault resilience of the W-cycle is retained for a 3D partial differential equation and higher order (quadratic) continuous finite elements, and

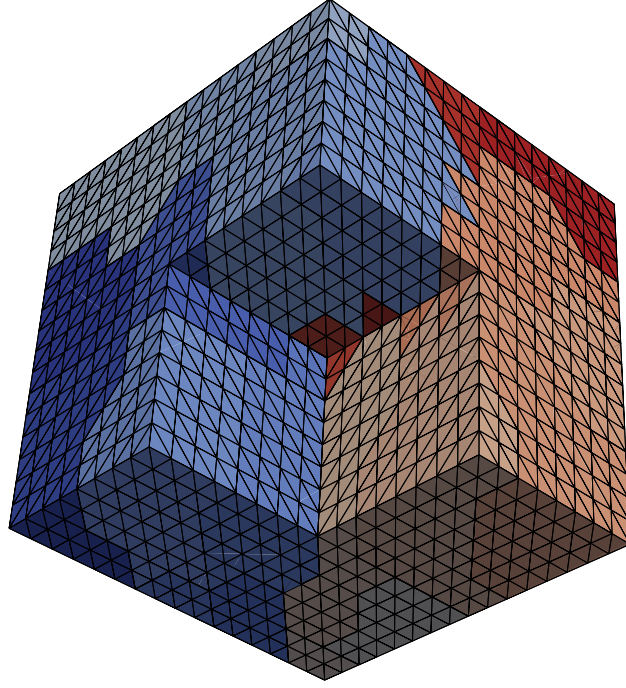


FIGURE 6. Mesh for the Fichera cube, indicating partitioning from METIS.

solve the Poisson equation on the Fichera cube.

$$\begin{aligned} -\Delta u &= f & \text{in } \Omega &= [0, 2]^3 \setminus [1, 2]^3, \\ u &= u_0 & \text{on } \partial\Omega \end{aligned}$$

The geometry and its uniform meshing are shown in Figure 6. The system is partitioned using METIS [14] into blocks of size 2^{14} and distributed over the compute nodes. We inject nodewise faults as given by (2). The problem sizes considered range from 1.7 million to about 940 million degrees of freedom. In Figure 6, we show the approximate rate of convergence as well as a level set of the error bound obtained for the Fault-Prone Two Grid Method in [1]. It can be seen that the rate of convergence in the multilevel case behaves like

$$\varrho(\mathcal{E}_L) = C_0 + \varepsilon \sqrt[n_L]{n_L},$$

where C_0 is a constant that is related to the fault-free method. We notice that the onset of divergent behaviour indeed happens at a slower rate as compared to the 2D setting. Protection of the prolongation again makes the method fault resilient, as shown in Figure 6. The choice of higher order elements plays no significant role in the convergence behaviour of the method.

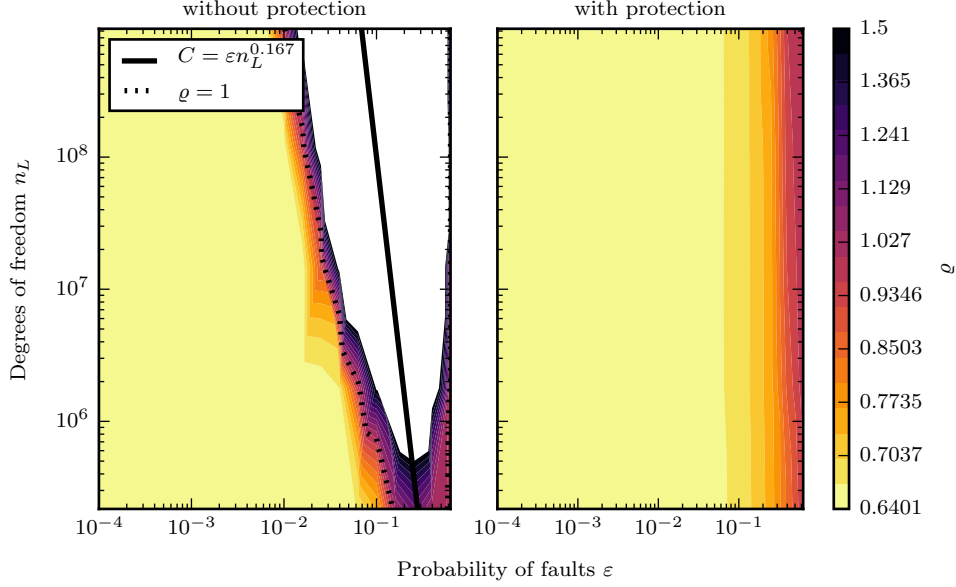


FIGURE 7. Lyapunov spectral radius $\rho(\mathcal{E}_L)$ for the iteration matrix for the Fault-Prone W-Cycle Multigrid Method in the case of discretization of the Fichera cube. Without protected prolongation (*left*), and with protected prolongation (*right*).

4. IMPLEMENTATION ISSUES

In the previous examples, we assumed that faults can be perfectly detected. In practice, faults cannot be perfectly detected, nor is it possible to perfectly protect the prolongation. The question therefore arises of whether Theorem 3 has any practical relevance. In Section 4.1, we present one possible simple approach for fault detection and show that the behaviour of Fault-Prone Multigrid found in (11) is recovered. In Section 4.2 we turn to the issue of protecting the prolongation operator.

4.1. Detection of Soft Faults. The laissez-faire fault mitigation strategy requires fault detection. While this is straight-forward in the case of hard faults, soft faults are more problematic. Various techniques have been suggested [16]. Here, we present a simple approach based on replication [13] in which a fault-prone component is repeated K times ($K \geq 2$) and the results are compared for consistency. The replication of node local operations is free of any communication requirements. Algorithm 2 shows how the strategy is used in conjunction with laissez-faire in the detection of faults in the computation of a generic matrix-vector product $y = \mathbf{M}x$. Instead of the action of \mathbf{M} , only its fault-prone equivalent \mathcal{M} is available in practice, where $\mathcal{M} \in \mathbb{R}^{n \times n}$ is a random matrix.

The basic idea behind Algorithm 2 is to declare an operation as being fault-free if all replicas are in agreement, otherwise a fault is deemed to have occurred and laissez-faire is triggered. This means that there are three possible outcomes for the

Algorithm 2 Detection of faults using K replicas and laissez-faire mitigation for the fault-prone matrix-vector product $y = \mathcal{M}x$.

```

1: for  $i \leftarrow 1$  to  $n$  do
2:   for  $j \leftarrow 1$  to  $K$  do
3:      $w_j \leftarrow (\mathcal{M}x)_i$  ▷  $j$ -th replica
4:   if  $w_1 = w_2 = \dots = w_K$  and  $|w_1| < 10^{16}$  then
5:      $y_i \leftarrow w_1$  ▷ value accepted
6:   else
7:      $y_i \leftarrow 0$  ▷ Laissez-faire

```

i -th component of the output y from Algorithm 2:

- (C) the correct value of $y_i = (\mathcal{M}x)_i$ is returned,
- (M) an error is detected, the laissez-faire mitigation is triggered, and $y_i = 0$,
- (U) a fault occurs but remains undetected, and y_i is a corrupted value.

Our analysis in Section 3 caters for the cases (C) and (M), but does not take account of case (U). Suppose that the probability of a replica being corrupted is $\varepsilon \ll 1$. Then the probability of *all* K replicas being corrupted in *exactly* the same way is $\mathcal{O}(\varepsilon^K)$, showing that the likelihood of (U) occurring is exponentially small in K . Hence, the probabilities of the outcomes of Algorithm 2 are

$$\begin{aligned}\mathbb{P}(C) &= 1 - \mathcal{O}(\varepsilon), \\ \mathbb{P}(M) &= \mathcal{O}(\varepsilon), \\ \mathbb{P}(U) &= \mathcal{O}(\varepsilon^K).\end{aligned}$$

We illustrate the effect of using our ad-hoc fault detection strategy given in Algorithm 2 in Figure 8 for the Poisson problem (10). In particular, the faults are modelled using bit flipping in which any bit in the floating point representation is flipped with a small but non-zero probability such that the overall probability of a floating point number being corrupted is ε . The results are given in Figure 8 in the simplest case where $K = 2$ replicas are used in Algorithm 2 and all operations. It is observed that, even in this simplest case, the convergence behaviour mirrors that which would be obtained with perfect detection.

4.2. Protection of the Prolongation. Theorem 3 requires that the prolongation operations

$$x_\ell \leftarrow x_\ell + \mathbf{P}e_{\ell-1}$$

are computed exactly, i.e. without *any* faults. This is clearly not possible in practice. We have seen that in the absence of full protection of the prolongation, the Fault-Prone Multigrid converges at a rate given by (11), where ε is the underlying failure rate of the machine. Moreover, Theorem 3 suggests that the $\mathcal{O}(\sqrt{n_L \varepsilon^2})$ term is entirely due to faults in the prolongation (with the faults coming purely from other components of multigrid contributing with higher order terms). At any rate, it is clear that if we can enhance the reliability of the prolongation *sufficiently* (without being necessarily exact), then one can expect to ameliorate the factor $\mathcal{O}(\sqrt{n_L \varepsilon^2})$ in the bound of the Lyapunov spectral radius, e.g. by obtaining a growth of $\mathcal{O}(n_L \varepsilon^\alpha)$ for some $\alpha > 2$.

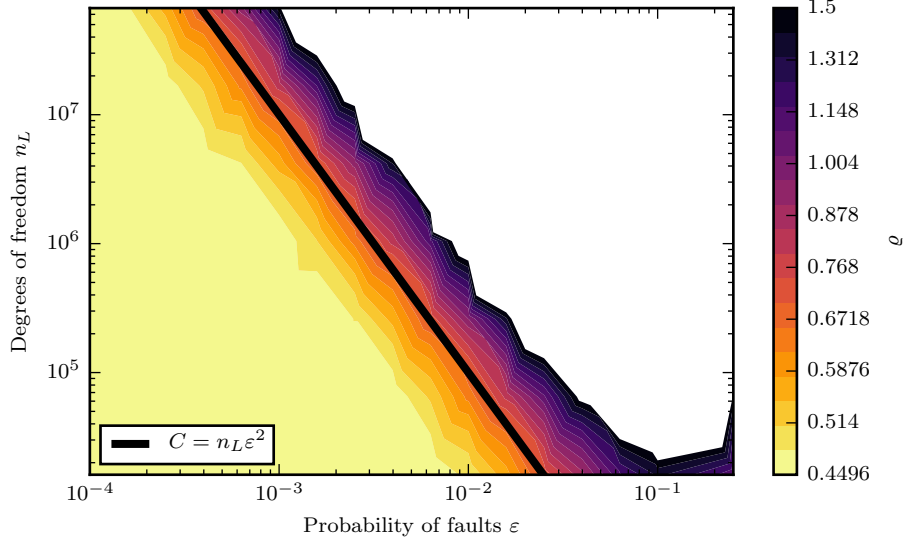


FIGURE 8. Lyapunov spectral radius $\varrho(\mathcal{E}_L)$ for the iteration matrix for the Fault-Prone W-Cycle Multigrid Method in the case of discretization of the 2D Poisson problem with fault detection using $K = 2$ replicas in every operation.

Is it possible to improve the likelihood of detecting and mitigating faults in the prolongation beyond $\mathbb{P}(M) = \mathcal{O}(\varepsilon)$? Algorithm 2 may be regarded as being overly conservative. For instance, if all but one of the replicas are in agreement, then intuitively it seems likely that the majority are correct. Algorithm 3 implements the approach suggested by this argument by looking for agreement amongst a subset of k_P of the K_P replicas for the prolongation. This added freedom alters the likelihood

Algorithm 3 Protection of the fault-prone prolongation $x_\ell \leftarrow x_\ell + \mathcal{P}e_{\ell-1}$ with up to K_P replicas, acceptance threshold k_P and laissez-faire mitigation.

```

1: for  $i \leftarrow 1$  to  $n_\ell$  do
2:   for  $j \leftarrow 1$  to  $K_P$  do
3:      $w_j \leftarrow (\mathcal{P}e_{\ell-1})_i$  ▷  $j$ -th replica
4:     if  $k_P$  replicas have matching value  $w$  and  $|w| < 10^{16}$  then
5:        $(x_\ell)_i \leftarrow (x_\ell)_i + w$  ▷ value accepted
6:       break

```

at which the three events occur in the prolongation:

$$\begin{aligned}\mathbb{P}(C_P) &= 1 - \mathcal{O}(\varepsilon^{K_P - k_P + 1}), \\ \mathbb{P}(M_P) &= \mathcal{O}(\varepsilon^{K_P - k_P + 1}), \\ \mathbb{P}(U_P) &= \mathcal{O}(\varepsilon^{k_P}).\end{aligned}$$

In particular, when Algorithm 3 is applied to the computation of the prolongation with parameters $k_P = 3$ and $K_P = 4$, we obtain an overall rate of $\mathbb{P}(M_P) =$

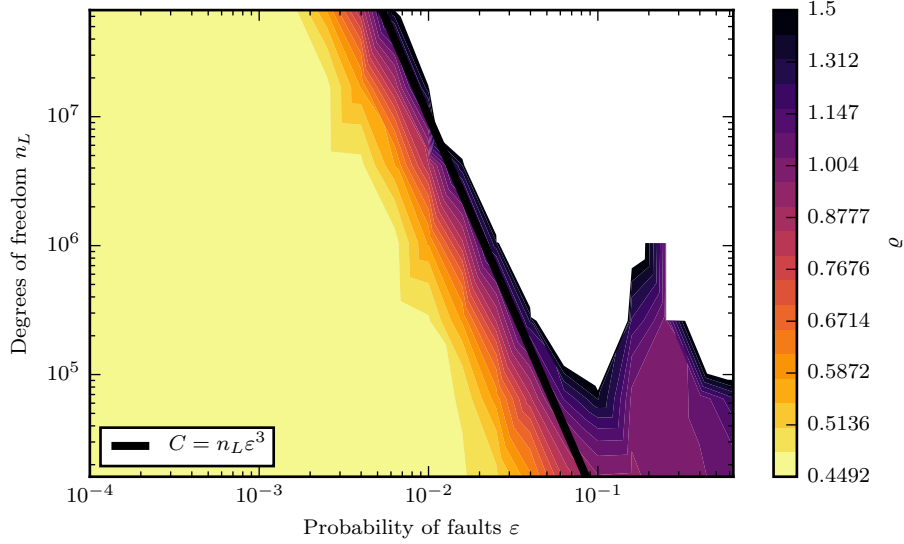


FIGURE 9. Lyapunov spectral radius $\rho(\mathcal{E}_L)$ for the iteration matrix for the Fault-Prone W-Cycle Multigrid Method with $K_P = 4$ replicas and acceptance threshold $k_P = 3$ in the prolongation and $K = 3$ replicas in all other operations.

$\mathcal{O}(\varepsilon^2)$ at which mitigation occurs. Figure 9 shows the results obtained by applying Algorithm 3 with $k_P = 3$ and $K_P = 4$ to the prolongation, and Algorithm 2 with $K = 3$ to all other operations. It is seen that the strategy has resulted in the factor $\mathcal{O}(\sqrt{n_L \varepsilon^2})$ in the Lyapunov spectral radius being replaced by $\mathcal{O}(n_L \varepsilon^\alpha)$ with $\alpha = 3$.

Would a cheaper detection and protection strategy (e.g. with $K_P = 3$, $k_P = 2$ and $K = 2$) suffice? Figure 10 shows the results obtained employing a variety of different choices for K , K_P and k_P . It is seen that the cheapest strategy that results in the growth $\mathcal{O}(\sqrt{n_L \varepsilon^2})$ being replaced by $\mathcal{O}(n_L \varepsilon^\alpha)$, $\alpha > 2$, is indeed $K_P = 4$, $k_P = 3$ and $K = 3$.

Moreover, if one wishes to obtain a given rate $\alpha \geq 2$, then one can choose $K = k_P = \alpha$ and $K_P = 2\alpha - 2$. As problems in three spatial dimensions are more resilient to the effect of laissez-faire mitigation, as reflected by (9) and the numerical evidence in Section 3, we expect the strategy $K = K_P = k_P = \alpha$ to be sufficient for $2 \leq \alpha \leq 6$.

5. CONCLUSION

In this work, we extended previous results concerning the convergence of the Fault-Prone Two Grid Method to the multigrid case, and showed that, if the prolongation is protected against faults, then the resulting Fault-Prone Multigrid Method is resilient. Numerical examples illustrated this result also holds in a variety of settings where the theory does not apply, along with the necessity of protecting the prolongation. Finally, we presented one possible simple fault detection strategy and showed that the resulting algorithm is also resilient.

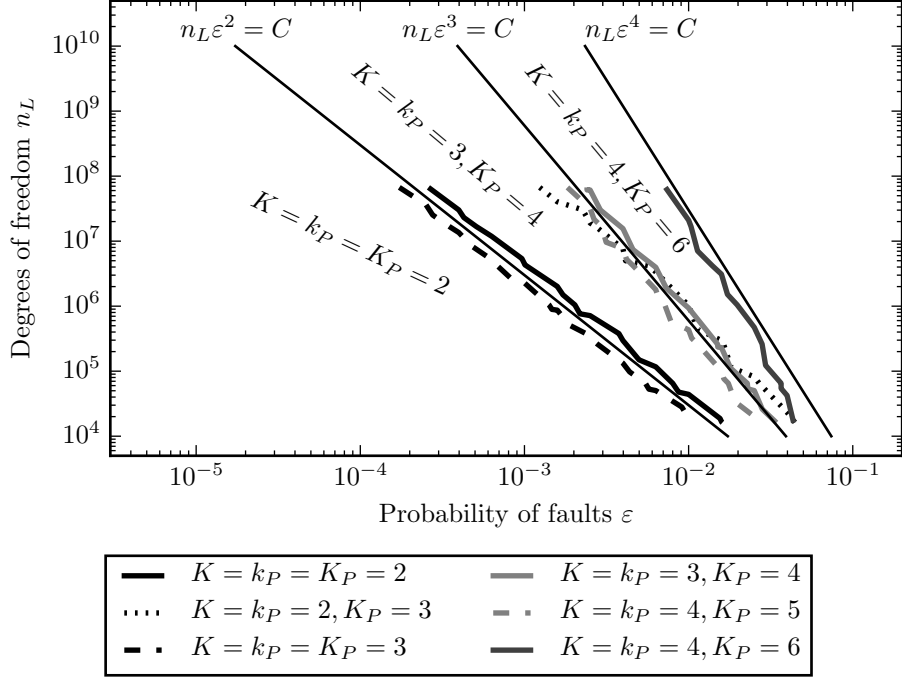


FIGURE 10. Level sets $\varrho(\mathcal{E}_L) = 0.57$ (50% more iterations than the ideal fault-free case) for different levels of detection and protection in the case of the 2D Poisson problem. Optimal parameter combinations are marked in thick solid lines. The obtained optimal detection and protection parameters for each region in the ε - n_L -plane are marked in the plot.

APPENDIX A. PROOF OF THEOREM 3

Throughout the appendices, C will be a generic constant whose value can change from line to line, but which is independent of ℓ and ε .

Definition 4 (Energy norms). *For matrices $\mathbf{Z} \in \mathbb{R}^{n_\ell \times n_k}$, we define the energy norm $\|\mathbf{Z}\|_A$ and the double energy norm $\|\mathbf{Z}\|_{A^2}$ as*

$$\|\mathbf{Z}\|_A = \left\| \mathbf{A}_\ell^{\frac{1}{2}} \mathbf{Z} \mathbf{A}_k^{-\frac{1}{2}} \right\|_2, \quad \|\mathbf{Z}\|_{A^2} = \left\| \mathbf{A}_\ell \mathbf{Z} \mathbf{A}_k^{-1} \right\|_2.$$

For matrices $\mathbf{W} \in \mathbb{R}^{n_\ell^2 \times n_k^2}$, we define the tensor energy norm $\|\mathbf{W}\|_A$ and the tensor double energy norm $\|\mathbf{W}\|_{A^2}$ as

$$\|\mathbf{W}\|_A = \left\| \left(\mathbf{A}_\ell^{\frac{1}{2}} \right)^{\otimes 2} \mathbf{W} \left(\mathbf{A}_k^{-\frac{1}{2}} \right)^{\otimes 2} \right\|_2, \quad \|\mathbf{W}\|_{A^2} = \left\| \mathbf{A}_\ell^{\otimes 2} \mathbf{W} \left(\mathbf{A}_k^{-1} \right)^{\otimes 2} \right\|_2.$$

In all cases $\|\bullet\|_2$ is the spectral norm.

We set

$$\mathbb{E} \left[\mathcal{X}_\ell^{(\bullet)} \right] = e_\ell^{(\bullet)} I, \quad \mathbb{V} \left[\mathcal{X}_\ell^{(\bullet)} \right] = \mathbf{V}_\ell^{(\bullet)}.$$

and expand

$$\begin{aligned}\mathbb{E} \left[\left(\boldsymbol{\varepsilon}_\ell^S \right)^{\otimes 2} \right] &= \mathbb{E} \left[\boldsymbol{\varepsilon}_\ell^S \right]^{\otimes 2} + \boldsymbol{C}_\ell^{(S)}, \\ \boldsymbol{C}_\ell^{(S)} &= \boldsymbol{V}_\ell^{(S)} (\boldsymbol{N}_\ell \boldsymbol{A}_\ell)^{\otimes 2},\end{aligned}$$

using Lemma 10 in [1]. Similar to the fault-free case, we derive recursive inequalities for the iteration matrix of the multigrid method.

Theorem 5. *If Assumptions 1 to 5 hold, then the iteration matrix of the Fault-Prone Multigrid Method satisfies the following recursive inequalities:*

(12)

$$\|\mathbb{E} [\boldsymbol{\varepsilon}_\ell(\nu_1, \nu_2, \gamma)]\|_{A^2} \leq \left\| \mathbb{E} [\boldsymbol{\varepsilon}_\ell^{TG}(\nu_1, \nu_2)] \right\|_{A^2} + C\varepsilon + C_* \|\mathbb{E} [\boldsymbol{\varepsilon}_{l-1}(\nu_1, \nu_2, \gamma)]\|_{A^2}^\gamma,$$

and

(13)

$$\begin{aligned}\left\| \mathbb{E} [\boldsymbol{\varepsilon}_\ell(\nu_1, \nu_2, \gamma)^{\otimes 2}] \right\|_{A^2} &\leq \left\| \mathbb{E} [\boldsymbol{\varepsilon}_\ell^{TG}(\nu_1, \nu_2)^{\otimes 2}] \right\|_{A^2} + C\varepsilon \\ &\quad + 2C_* \left\| \mathbb{E} [\boldsymbol{\varepsilon}_\ell^{TG}(\nu_1, \nu_2)] \right\|_{A^2} \|\mathbb{E} [\boldsymbol{\varepsilon}_{l-1}(\nu_1, \nu_2, \gamma)]\|_{A^2}^\gamma \\ &\quad + C_*^2 \left\| \mathbb{E} [\boldsymbol{\varepsilon}_{l-1}(\nu_1, \nu_2, \gamma)^{\otimes 2}] \right\|_{A^2}^\gamma,\end{aligned}$$

with

$$C_* = C_S \underline{C}_p \overline{C}_p (C_S + C_A \eta(\nu_2)).$$

The constants C and C_* are independent of the level ℓ and ε .

If $\varepsilon = 0$, the second inequality matches exactly the recursive inequality that we would use for the fault-free case, while the first inequality matches the tensor square power of the same inequality.

Proof. Since the proof on level ℓ only involves the levels ℓ and $\ell - 1$, we will drop the first subscript and replace the second one with a subscript C .

We can write the iteration matrix as

$$\begin{aligned}\boldsymbol{\varepsilon}(\nu_1, \nu_2, \gamma) &= \left(\boldsymbol{\varepsilon}^{S, \text{post}} \right)^{\nu_2} \left(\boldsymbol{I} - \boldsymbol{P} (\boldsymbol{I} - (\boldsymbol{\varepsilon}_C)^\gamma) \boldsymbol{A}_C^{-1} \boldsymbol{\chi}_C^{(R)} \boldsymbol{R} \boldsymbol{\chi}^{(\rho)} \boldsymbol{A} \right) \left(\boldsymbol{\varepsilon}^{S, \text{pre}} \right)^{\nu_1} \\ &= \underbrace{\left(\boldsymbol{\varepsilon}^{S, \text{post}} \right)^{\nu_2} \left(\boldsymbol{I} - \boldsymbol{P} \boldsymbol{A}_C^{-1} \boldsymbol{\chi}_C^{(R)} \boldsymbol{R} \boldsymbol{\chi}^{(\rho)} \boldsymbol{A} \right) \left(\boldsymbol{\varepsilon}^{S, \text{pre}} \right)^{\nu_1}}_{\boldsymbol{\varepsilon}^{TG}(\nu_1, \nu_2)} \\ &\quad + \underbrace{\left(\boldsymbol{\varepsilon}^{S, \text{post}} \right)^{\nu_2} \boldsymbol{P} (\boldsymbol{\varepsilon}_C)^\gamma \boldsymbol{A}_C^{-1} \boldsymbol{\chi}_C^{(R)} \boldsymbol{R} \boldsymbol{\chi}^{(\rho)} \boldsymbol{A} \left(\boldsymbol{\varepsilon}^{S, \text{pre}} \right)^{\nu_1}}_{=: \boldsymbol{C}}.\end{aligned}$$

Since

$$\begin{aligned}\left\| \mathbb{E} \left[\left(\boldsymbol{\varepsilon}^S \right)^{\nu_2} \boldsymbol{P} \right] \right\|_{A^2} &= \left\| \boldsymbol{A} \mathbb{E} \left[\boldsymbol{\varepsilon}^S \right]^{\nu_2} \boldsymbol{P} \boldsymbol{A}_C^{-1} \right\|_2 \\ &= \left\| \boldsymbol{A}_C^{-1} \boldsymbol{R} \boldsymbol{A} \mathbb{E} \left[\boldsymbol{\varepsilon}^S \right]^{\nu_2} \right\|_2 \\ &\leq \underline{C}_p \left(\left\| \mathbb{E} \left[\boldsymbol{\varepsilon}^S \right]^{\nu_2} \right\|_2 + \left\| \boldsymbol{E}^{CG} \mathbb{E} \left[\boldsymbol{\varepsilon}^S \right]^{\nu_2} \right\|_2 \right) \\ &\leq \underline{C}_p (C_S + C_A \eta(\nu_2)) + C\varepsilon,\end{aligned}$$

$$\begin{aligned} \left\| \mathbb{E} \left[\mathbf{A}_C^{-1} \boldsymbol{\chi}_C^{(R)} \mathbf{R} \boldsymbol{\chi}^{(\rho)} \mathbf{A} \right] \right\|_{A^2} &\leq \left| e_C^{(R)} e^{(\rho)} \right| \|\mathbf{R}\|_2 \leq \overline{C}_p + C\varepsilon, \\ \left\| \mathbb{E} \left[\boldsymbol{\varepsilon}^S \right]^{\nu_1} \right\|_{A^2} &\leq C_S + C\varepsilon, \end{aligned}$$

we have by Assumptions 1 and 5 that

$$\begin{aligned} \|\mathbb{E}[\mathbf{C}]\|_{A^2} &\leq \left\| \mathbb{E} \left[\boldsymbol{\varepsilon}^S \right]^{\nu_2} \mathbf{P} \right\|_{A^2} \|\mathbb{E}[\boldsymbol{\varepsilon}_C]\|_{A^2}^\gamma \left\| \mathbb{E} \left[\mathbf{A}_C^{-1} \boldsymbol{\chi}_C^{(R)} \mathbf{R} \boldsymbol{\chi}^{(\rho)} \mathbf{A} \right] \right\|_{A^2} \left\| \mathbb{E} \left[\boldsymbol{\varepsilon}^S \right]^{\nu_1} \right\|_{A^2} \\ (14) \quad &\leq C_* \|\mathbb{E}[\boldsymbol{\varepsilon}_C]\|_{A^2}^\gamma + C\varepsilon. \end{aligned}$$

Therefore the first inequality of the theorem follows from

$$\|\mathbb{E}[\boldsymbol{\varepsilon}]\|_{A^2} \leq \left\| \mathbb{E} \left[\boldsymbol{\varepsilon}^{TG} \right] \right\|_{A^2} + \|\mathbb{E}[\mathbf{C}]\|_{A^2}.$$

The second inequality is more work, but follows the same path:

$$(15) \quad \|\mathbb{E}[\boldsymbol{\varepsilon}^{\otimes 2}]\|_{A^2} \leq \left\| \mathbb{E} \left[\left(\boldsymbol{\varepsilon}^{TG} \right)^{\otimes 2} \right] \right\|_{A^2} + \|\mathbb{E}[\mathbf{C}^{\otimes 2}]\|_{A^2} + 2 \left\| \mathbb{E} \left[\boldsymbol{\varepsilon}^{TG} \otimes \mathbf{C} \right] \right\|_{A^2}.$$

We start by bounding the second term on the right-hand side.

$$\begin{aligned} (16) \quad \|\mathbb{E}[\mathbf{C}^{\otimes 2}]\|_{A^2} &\leq \left\| \mathbb{E} \left[\left(\left(\boldsymbol{\varepsilon}^S \right)^{\nu_2} \mathbf{P} \right)^{\otimes 2} \right] \right\|_{A^2} \|\mathbb{E}[\boldsymbol{\varepsilon}_C^{\otimes 2}]\|_{A^2}^\gamma \\ &\quad \times \left\| \mathbb{E} \left[\left(\mathbf{A}_C^{-1} \boldsymbol{\chi}_C^{(R)} \mathbf{R} \boldsymbol{\chi}^{(\rho)} \mathbf{A} \right)^{\otimes 2} \right] \right\|_{A^2} \left\| \mathbb{E} \left[\left(\boldsymbol{\varepsilon}^S \right)^{\nu_1} \right]^{\otimes 2} \right\|_{A^2} \end{aligned}$$

We have by

$$\begin{aligned} (17) \quad \left\| \mathbb{E} \left[\left(\left(\boldsymbol{\varepsilon}^S \right)^{\nu_2} \mathbf{P} \right)^{\otimes 2} \right] \right\|_{A^2} &\leq \left\| \left(\mathbb{E} \left[\boldsymbol{\varepsilon}^S \right]^{\nu_2} \mathbf{P} \right)^{\otimes 2} \right\|_{A^2} + C \|\mathbf{C}^{(S)}\|_{A^2} \\ &= \left\| \mathbf{A} \mathbb{E} \left[\boldsymbol{\varepsilon}^S \right]^{\nu_2} \mathbf{P} \mathbf{A}_C^{-1} \right\|_2^2 + C\varepsilon \\ &= \left\| \mathbf{A}_C^{-1} \mathbf{R} \mathbf{A} \mathbb{E} \left[\boldsymbol{\varepsilon}^S \right]^{\nu_2} \right\|_2^2 + C\varepsilon \\ &\leq \underline{C}_p^2 \left(\left\| \mathbb{E} \left[\boldsymbol{\varepsilon}^S \right]^{\nu_2} \right\|_2 + \left\| \mathbf{E}^{CG} \mathbb{E} \left[\boldsymbol{\varepsilon}^S \right]^{\nu_2} \right\|_2 \right)^2 + C\varepsilon \\ &\leq \underline{C}_p^2 (C_S + C_A \eta(\nu_2))^2 + C\varepsilon. \end{aligned}$$

Since by Assumptions 1 and 5

$$\begin{aligned} (18) \quad \left\| \mathbb{E} \left[\left(\mathbf{A}_C^{-1} \boldsymbol{\chi}_C^{(R)} \mathbf{R} \boldsymbol{\chi}^{(\rho)} \mathbf{A} \right)^{\otimes 2} \right] \right\|_{A^2} &= \left\| \mathbb{E} \left[\left(\boldsymbol{\chi}_C^{(R)} \mathbf{R} \boldsymbol{\chi}^{(\rho)} \right)^{\otimes 2} \right] \right\|_2 \\ &\leq \left\| \mathbb{E} \left[\left(\boldsymbol{\chi}_C^{(R)} \right)^{\otimes 2} \right] \right\|_2 \|\mathbf{R}\|_2^2 \left\| \mathbb{E} \left[\left(\boldsymbol{\chi}^{(\rho)} \right)^{\otimes 2} \right] \right\|_2 \\ &\leq \left(\left\| \mathbb{V} \left[\boldsymbol{\chi}_C^{(R)} \right] \right\|_2 + \left\| \mathbb{E} \left[\boldsymbol{\chi}_C^{(R)} \right] \right\|_2^2 \right) \|\mathbf{R}\|_2^2 \\ &\quad \times \left(\left\| \mathbb{V} \left[\boldsymbol{\chi}^{(\rho)} \right] \right\|_2 + \left\| \mathbb{E} \left[\boldsymbol{\chi}^{(\rho)} \right] \right\|_2^2 \right) \\ &\leq \left(\varepsilon + \left(e^{(R)} \right)^2 \right) \left(\varepsilon + \left(e^{(\rho)} \right)^2 \right) \|\mathbf{R}\|_2^2 \\ &\leq \overline{C}_p^2 + C\varepsilon. \end{aligned}$$

and

$$(19) \quad \left\| \mathbb{E} \left[\left((\mathcal{E}^S)^{\nu_1} \right)^{\otimes 2} \right] \right\|_{A^2} = \left\| \left(\mathbb{E} [\mathcal{E}^S]^{\otimes 2} + \mathcal{C}^{(S)} \right)^{\nu_1} \right\|_{A^2} \\ \leq \left\| \mathbb{E} [\mathcal{E}^S]^{\nu_1} \right\|_{A^2}^2 + C\varepsilon \leq C_S^2 + C\varepsilon,$$

Inserting eqs. (17), (18) and (19) into (16), we have

$$(20) \quad \left\| \mathbb{E} [\mathcal{C}^{\otimes 2}] \right\|_{A^2} \leq C_*^2 \left\| \mathbb{E} [\mathcal{E}_C^{\otimes 2}] \right\|_{A^2}^\gamma + C\varepsilon.$$

Now we turn to the third term of (15). We split

$$\begin{aligned} \mathcal{E}^{TG} &= \left(\mathcal{E}^{S,\text{post}} \right)^{\nu_2} \left(I - \mathbf{P} \mathbf{A}_C^{-1} \mathcal{X}_C^{(R)} \mathbf{R} \mathcal{X}^{(\rho)} \mathbf{A} \right) \left(\mathcal{E}^{S,\text{pre}} \right)^{\nu_1} \\ &= \mathbb{E} [\mathcal{E}^S]^{\nu_2} \mathbf{E}^{CG} \mathbb{E} [\mathcal{E}^S]^{\nu_1} \\ &\quad + \left(\left(\mathcal{E}^{S,\text{post}} \right)^{\nu_2} - \mathbb{E} [\mathcal{E}^S]^{\nu_2} \right) \mathbf{E}^{CG} \mathbb{E} [\mathcal{E}^S]^{\nu_1} \\ &\quad + \mathbb{E} [\mathcal{E}^S]^{\nu_2} \mathbf{E}^{CG} \left(\left(\mathcal{E}^{S,\text{pre}} \right)^{\nu_1} - \mathbb{E} [\mathcal{E}^S]^{\nu_1} \right) \\ &\quad + \left(\left(\mathcal{E}^{S,\text{post}} \right)^{\nu_2} - \mathbb{E} [\mathcal{E}^S]^{\nu_2} \right) \mathbf{E}^{CG} \left(\left(\mathcal{E}^{S,\text{pre}} \right)^{\nu_1} - \mathbb{E} [\mathcal{E}^S]^{\nu_1} \right) \\ &\quad + \left(\mathcal{E}^{S,\text{post}} \right)^{\nu_2} \mathbf{P} \left(\mathbf{A}_C^{-1} \mathbf{R} \mathbf{A} - \mathbf{A}_C^{-1} \mathcal{X}_C^{(R)} \mathbf{R} \mathcal{X}^{(\rho)} \mathbf{A} \right) \left(\mathcal{E}^{S,\text{pre}} \right)^{\nu_1} \end{aligned}$$

so that

$$(21) \quad \left\| \mathbb{E} [\mathcal{E}^{TG} \otimes \mathcal{C}] \right\|_{A^2} \\ \leq \left\| \mathbb{E} [\mathcal{E}^S]^{\nu_2} \mathbf{E}^{CG} \mathbb{E} [\mathcal{E}^S]^{\nu_1} \right\|_{A^2} \left\| \mathbb{E} [\mathcal{C}] \right\|_{A^2} \\ + \left\| \mathbf{E}^{CG} \right\|_{A^2} \left\| \mathbf{P} \right\|_{A^2} \left\| \mathbb{E} [\mathcal{E}_C] \right\|_{A^2}^\gamma \left\| \mathbb{E} [\mathbf{A}_C^{-1} \mathcal{X}_C^{(R)} \mathbf{R} \mathcal{X}^{(\rho)} \mathbf{A}] \right\|_{A^2} \\ \left\{ \left\| \mathbb{E} \left[\left(\mathcal{E}^S \right)^{\nu_2} \otimes \left(\left(\mathcal{E}^S \right)^{\nu_2} - \mathbb{E} [\mathcal{E}^S]^{\nu_2} \right) \right] \right\|_{A^2} \left\| \mathbb{E} [\mathcal{E}^S]^{\nu_1} \right\|_{A^2}^2 \right. \\ + \left\| \mathbb{E} [\mathcal{E}^S]^{\nu_2} \right\|_{A^2}^2 \left\| \mathbb{E} \left[\left(\mathcal{E}^S \right)^{\nu_1} \otimes \left(\left(\mathcal{E}^S \right)^{\nu_1} - \mathbb{E} [\mathcal{E}^S]^{\nu_1} \right) \right] \right\|_{A^2} \\ + \left\| \mathbb{E} \left[\left(\mathcal{E}^S \right)^{\nu_2} \otimes \left(\left(\mathcal{E}^S \right)^{\nu_2} - \mathbb{E} [\mathcal{E}^S]^{\nu_2} \right) \right] \right\|_{A^2} \\ \times \left\| \mathbb{E} \left[\left(\mathcal{E}^S \right)^{\nu_1} \otimes \left(\left(\mathcal{E}^S \right)^{\nu_1} - \mathbb{E} [\mathcal{E}^S]^{\nu_1} \right) \right] \right\|_{A^2} \left. \right\} \\ + \left\| \mathbb{E} \left[\left(\mathbf{A}_C^{-1} \mathbf{R} \mathbf{A} \right) \otimes \left(\mathbf{A}_C^{-1} \mathcal{X}_C^{(R)} \mathbf{R} \mathcal{X}^{(\rho)} \mathbf{A} \right) - \left(\mathbf{A}_C^{-1} \mathcal{X}_C^{(R)} \mathbf{R} \mathcal{X}^{(\rho)} \mathbf{A} \right)^{\otimes 2} \right] \right\|_{A^2} \\ \times \left\| \mathbb{E} [\mathcal{E}_C] \right\|_{A^2}^\gamma \left\| \mathbb{E} \left[\left(\left(\mathcal{E}^S \right)^{\nu} \mathbf{P} \right)^{\otimes 2} \right] \right\|_{A^2} \left\| \mathbb{E} \left[\left(\left(\mathcal{E}^S \right)^{\nu_1} \right)^{\otimes 2} \right] \right\|_{A^2}$$

Now, we have by Assumptions 3 and 5 that

$$(22) \quad \left\| \mathbf{E}^{CG} \right\|_{A^2} = \left\| \mathbf{E}^{CG} \right\|_2 \leq C_A,$$

$$(23) \quad \left\| \mathbf{P} \right\|_{A^2} = \left\| \mathbf{A} \mathbf{P} \mathbf{A}_C^{-1} \right\|_2 \leq \underline{C}_p,$$

and by Assumptions 1 and 5 that

$$(24) \quad \left\| \mathbb{E} \left[\left(\mathbf{A}_C^{-1} \mathbf{R} \mathbf{A} \right) \otimes \left(\mathbf{A}_C^{-1} \mathcal{X}_C^{(R)} \mathbf{R} \mathcal{X}^{(\rho)} \mathbf{A} \right) - \left(\mathbf{A}_C^{-1} \mathcal{X}_C^{(R)} \mathbf{R} \mathcal{X}^{(\rho)} \mathbf{A} \right)^{\otimes 2} \right] \right\|_{A^2}$$

$$\begin{aligned}
&= \left\| \mathbb{E} \left[\mathbf{R} \otimes \left(\boldsymbol{\chi}_C^{(R)} \mathbf{R} \boldsymbol{\chi}^{(\rho)} \right) - \left(\boldsymbol{\chi}_C^{(R)} \mathbf{R} \boldsymbol{\chi}^{(\rho)} \right)^{\otimes 2} \right] \right\|_2 \\
&\leq \|\mathbf{R}\|_2^2 \left\{ \left| e_C^{(R)} e^{(\rho)} \right| \left| 1 - e_C^{(R)} e^{(\rho)} \right| + \left(e^{(\rho)} \right)^2 \left\| \mathbb{V} \left[\boldsymbol{\chi}_C^{(R)} \right] \right\|_2 \right. \\
&\quad \left. + \left(e_C^{(R)} \right)^2 \left\| \mathbb{V} \left[\boldsymbol{\chi}^{(\rho)} \right] \right\|_2 + \left\| \mathbb{V} \left[\boldsymbol{\chi}_C^{(R)} \right] \right\|_2 \left\| \mathbb{V} \left[\boldsymbol{\chi}^{(\rho)} \right] \right\|_2 \right\} \\
&\leq C\varepsilon
\end{aligned}$$

and that

$$\begin{aligned}
(25) \quad &\left\| \mathbb{E} \left[\left(\boldsymbol{\varepsilon}^S \right)^\nu \otimes \left(\left(\boldsymbol{\varepsilon}^S \right)^\nu - \mathbb{E} \left[\boldsymbol{\varepsilon}^S \right]^\nu \right) \right] \right\|_{A^2} \\
&= \left\| \mathbb{E} \left[\left(\boldsymbol{\varepsilon}^S \right)^{\otimes 2} \right]^\nu - \left(\mathbb{E} \left[\boldsymbol{\varepsilon}^S \right]^{\otimes 2} \right)^\nu \right\|_{A^2} \\
&= \left\| \left(\mathbb{E} \left[\boldsymbol{\varepsilon}^S \right]^{\otimes 2} + \mathbf{C}^{(S)} \right)^\nu - \left(\mathbb{E} \left[\boldsymbol{\varepsilon}^S \right]^{\otimes 2} \right)^\nu \right\|_{A^2} \\
&\leq C \left\| \mathbf{C}^{(S)} \right\|_{A^2} \leq C\varepsilon.
\end{aligned}$$

Also,

$$\begin{aligned}
(26) \quad &\left\| \mathbb{E} \left[\boldsymbol{\varepsilon}^S \right]^{\nu_2} \mathbf{E}^{CG} \mathbb{E} \left[\boldsymbol{\varepsilon}^S \right]^{\nu_1} \right\|_{A^2} \\
&\leq \left\| \mathbb{E} \left[\boldsymbol{\varepsilon}^{TG}(\nu_1, \nu_2) \right] \right\|_{A^2} + \left| 1 - e_C^{(R)} \right| \left| 1 - e^{(\rho)} \right| \left\| \mathbb{E} \left[\boldsymbol{\varepsilon}^S \right]^{\nu_1} (\mathbf{I} - \mathbf{E}^{CG}) \mathbb{E} \left[\boldsymbol{\varepsilon}^S \right]^{\nu_2} \right\|_2 \\
&= \left\| \mathbb{E} \left[\boldsymbol{\varepsilon}^{TG}(\nu_1, \nu_2) \right] \right\|_{A^2} + C\varepsilon.
\end{aligned}$$

Hence we find by inserting eqs. (24) to (25), (14), (22) and (23) into (21)

$$(27) \quad \left\| \mathbb{E} \left[\boldsymbol{\varepsilon}^{TG} \otimes \mathbf{C} \right] \right\|_{A^2} \leq C_* \left\| \mathbb{E} \left[\boldsymbol{\varepsilon}^{TG} \right] \right\|_{A^2} \left\| \mathbb{E} \left[\boldsymbol{\varepsilon}_C \right] \right\|_{A^2}^\gamma + C\varepsilon.$$

Therefore, the second recursive inequality

$$\begin{aligned}
\left\| \mathbb{E} \left[\boldsymbol{\varepsilon}^{\otimes 2} \right] \right\|_{A^2} &\leq \left\| \mathbb{E} \left[\left(\boldsymbol{\varepsilon}^{TG} \right)^{\otimes 2} \right] \right\|_{A^2} + C\varepsilon \\
&\quad + 2C_* \left\| \mathbb{E} \left[\boldsymbol{\varepsilon}^{TG} \right] \right\|_{A^2} \left\| \mathbb{E} \left[\boldsymbol{\varepsilon}_C \right] \right\|_{A^2}^\gamma \\
&\quad + (C_*)^2 \left\| \mathbb{E} \left[\boldsymbol{\varepsilon}_C^{\otimes 2} \right] \right\|_{A^2}^\gamma
\end{aligned}$$

follows from eqs. (15), (20) and (27). \square

We adapt Theorem 1 to the tensor space setting:

Lemma 6. *Let the elements of the sequences $\{\eta_\ell\}_{\ell \geq 1}$ and $\{\eta_{\ell, \otimes 2}\}_{\ell \geq 1}$ be bounded as follows*

$$\begin{aligned}
\eta_1 &\leq \xi, & \eta_\ell &\leq \xi + C_* \eta_{\ell-1}^\gamma, & \ell &\geq 2, \\
\eta_{1, \otimes 2}^2 &\leq \xi_{\otimes 2}^2, & \eta_{\ell, \otimes 2}^2 &\leq \xi_{\otimes 2}^2 + 2\xi C_* \eta_{\ell-1}^\gamma + C_*^2 \eta_{\ell-1, \otimes 2}^{2\gamma}, & \ell &\geq 2.
\end{aligned}$$

If $\gamma \geq 2$, $C_* \gamma > 1$ and $\max \{\xi, \xi_{\otimes 2}\} \leq \frac{\gamma-1}{\gamma} \frac{1}{(C_* \gamma)^{\frac{1}{\gamma-1}}}$, then

$$\max \{\eta_\ell, \eta_{\ell, \otimes 2}\} \leq \frac{\gamma}{\gamma-1} \max \{\xi, \xi_{\otimes 2}\} < 1, \quad \ell \geq 1.$$

In the case of $\gamma = 2$, we have

$$\max \{\eta_\ell, \eta_{\ell, \otimes 2}\} \leq \frac{2 \max \{\xi, \xi_{\otimes 2}\}}{1 + \sqrt{1 - 4C_* \max \{\xi, \xi_{\otimes 2}\}}} < 1, \quad \ell \geq 1.$$

Proof. Set $\alpha_\ell := \max \{\eta_\ell, \eta_{\ell, \otimes 2}\}$ and $\beta := \max \{\xi, \xi_{\otimes 2}\}$. Then

$$\begin{aligned} \eta_1 &\leq \beta, & \eta_\ell &\leq \beta + C_* \alpha_{\ell-1}^\gamma, \quad \ell \geq 2, \\ \eta_{1, \otimes 2}^2 &\leq \beta^2, & \eta_{\ell, \otimes 2}^2 &\leq \beta^2 + 2\beta C_* \alpha_{\ell-1}^\gamma + C_*^2 \alpha_{\ell-1}^{2\gamma}, \quad \ell \geq 2, \end{aligned}$$

and hence

$$\alpha_1 \leq \beta, \quad \alpha_\ell \leq \beta + C_* \alpha_{\ell-1}^\gamma, \quad \ell \geq 2.$$

Application of theorem 1 gives for $\gamma \geq 2$

$$\alpha_\ell \leq \frac{\gamma}{\gamma-1} \beta < 1$$

and for $\gamma = 2$

$$\alpha_\ell \leq \frac{2\beta}{1 + \sqrt{1 - 4C_*\beta}} < 1.$$

□

Theorem 3. *Provided Assumptions 1 to 5 hold, that the prolongation is protected, that the number of smoothing steps is sufficient, and that ε is sufficiently small, the fault-prone multigrid method converges with a rate bounded by*

$$\varrho(\mathcal{E}_\ell(\nu_1, \nu_2, \gamma)) \leq \begin{cases} \frac{\gamma}{\gamma-1} C_{TG} + C\varepsilon, & \gamma \geq 2, \\ \frac{2}{1 + \sqrt{1 - 4C_* C_{TG}}} C_{TG} + C\varepsilon, & \gamma = 2, \end{cases}$$

where C is independent of ε and ℓ and

$$\begin{aligned} C_{TG} &= \max_\ell \left\| \mathbf{E}_\ell^{TG}(\nu_2, \nu_1) \right\|_2 \leq C_S C_A \eta(\nu_2), \\ C_* &= C_S \underline{C}_p \overline{C}_p (C_S + C_A \eta(\nu_2)). \end{aligned}$$

Proof. Set

$$\begin{aligned} \eta_\ell &:= \left\| \mathbb{E}[\mathcal{E}_\ell(\nu_1, \nu_2, \gamma)] \right\|_{A^2}, \\ \eta_{\ell, \otimes 2} &:= \left\| \mathbb{E}[\mathcal{E}_\ell(\nu_1, \nu_2, \gamma)^{\otimes 2}] \right\|_{A^2}^{\frac{1}{2}}, \\ \xi &:= \max_\ell \left\| \mathbf{E}_\ell^{TG}(\nu_2, \nu_1) \right\|_2 + C\varepsilon, \\ \xi_{\otimes 2} &:= \max_\ell \left\| \mathbf{E}_\ell^{TG}(\nu_2, \nu_1) \right\|_2 + C\varepsilon. \end{aligned}$$

It follows from

$$\begin{aligned} \mathbb{E}[\mathcal{E}_\ell^S] &= \mathbf{E}_\ell^S + \left(1 - e_\ell^{(S)}\right) (\mathbf{I} - \mathbf{E}_\ell^S), \\ \mathbb{E}[\mathcal{E}_\ell^{CG}] &= \mathbf{E}_\ell^{CG} + \left(1 - e_{\ell-1}^{(R)} e_\ell^{(\rho)}\right) (\mathbf{I} - \mathbf{E}_\ell^{CG}), \end{aligned}$$

and Assumptions 1, 3 and 4 that

$$\left\| \mathbb{E}[\mathcal{E}_\ell^{TG}(\nu_1, \nu_2)] \right\|_{A^2} \leq \left\| \mathbf{E}_\ell^{TG}(\nu_2, \nu_1) \right\|_2 + C\varepsilon,$$

and therefore with (12) that $\eta_\ell \leq \xi + C_* \eta_{\ell-1}^\gamma$. By Theorem 2 and (13), we find $\eta_{\ell, \otimes 2}^2 \leq \xi_{\otimes 2}^2 + 2\xi C_* \eta_{\ell-1}^\gamma + C_*^2 \eta_{\ell-1, \otimes 2}^2$. Since C_* , ξ and $\xi_{\otimes 2}$ decrease as the number of smoothing steps increases, Theorem 6 can be applied for a sufficient number of smoothing steps. \square

ACKNOWLEDGMENTS

This research used resources of the Oak Ridge Leadership Computing Facility at the Oak Ridge National Laboratory, which is supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC05-00OR22725.

REFERENCES

- [1] M. AINSWORTH AND C. GLUSA, *Is the multigrid method fault tolerant? The Two Grid Case.*, SIAM Journal on Scientific Computing, (Submitted).
- [2] A. AVIZIENIS, J.-C. LAPRIE, B. RANDELL, AND C. LANDWEHR, *Basic concepts and taxonomy of dependable and secure computing*, IEEE Transactions on Dependable and Secure Computing, 1 (2004), pp. 11–33.
- [3] R. E. BANK, A. H. SHERMAN, AND A. WEISER, *Some refinement algorithms and data structures for regular local mesh refinement*, Scientific Computing, Applications of Mathematics and Computing to the Physical Sciences, 1 (1983), pp. 3–17.
- [4] P. BOUGEROL AND J. LACROIX, *Products of random matrices with applications to Schrödinger operators*, vol. 8 of Progress in Probability and Statistics, Birkhäuser Boston Inc., Boston, MA, 1985.
- [5] D. BRAESS, *Finite elements. Theory, fast solvers and applications in solid mechanics. Translated from German by Larry L. Schumaker. 3rd ed.*, Cambridge: Cambridge University Press, 2007, <http://dx.doi.org/10.1017/CB09780511618635>.
- [6] F. CAPPELLO, A. GEIST, B. GROPP, L. KALE, B. KRAMER, AND M. SNIR, *Toward exascale resilience*, International Journal of High Performance Computing Applications, (2009).
- [7] F. CAPPELLO, A. GEIST, W. GROPP, S. KALE, B. KRAMER, AND M. SNIR, *Toward exascale resilience: 2014 update*, Supercomputing frontiers and innovations, 1 (2014), pp. 5–28.
- [8] A. CRISANTI, G. PALADIN, AND A. VULPIANI, *Products of random matrices*, Springer, 1993.
- [9] W. DÖRFLER, *A convergent adaptive algorithm for Poisson’s equation*, SIAM Journal on Numerical Analysis, 33 (1996), pp. 1106–1124.
- [10] A. ERN AND J.-L. GUERMOND, *Theory and Practice of Finite Elements.*, Applied Mathematical Sciences 159. New York, NY: Springer, 2004.
- [11] W. HACKBUSCH, *Multi-grid methods and applications*, vol. 4, Springer-Verlag Berlin, 1985.
- [12] W. HACKBUSCH, *Iterative solution of large sparse systems of equations*, vol. 95 of Applied Mathematical Sciences, Springer-Verlag, New York, 1994, <http://dx.doi.org/10.1007/978-1-4612-4288-8>.
- [13] T. HERAULT AND Y. ROBERT, *Fault-Tolerance Techniques for High-Performance Computing*, Springer, 2015.
- [14] G. KARYPIS AND V. KUMAR, *A fast and high quality multilevel scheme for partitioning irregular graphs*, SIAM Journal on Scientific Computing, 20 (1998), pp. 359–392, <http://dx.doi.org/10.1137/S1064827595287997>.
- [15] T. KRÖGER AND T. PREUSSER, *Stability of the 8-tetrahedra shortest-interior-edge partitioning method*, Numerische Mathematik, 109 (2008), pp. 435–457.
- [16] M. SNIR, R. W. WISNIEWSKI, J. A. ABRAHAM, S. V. ADVE, S. BAGCHI, P. BALAJI, J. BELAK, P. BOSE, F. CAPPELLO, B. CARLSON, ET AL., *Addressing failures in exascale computing*, International Journal of High Performance Computing Applications, 28 (2014), pp. 129–173.
- [17] U. TROTTEMBERG, C. W. OOSTERLEE, AND A. SCHÜLLER, *Multigrid*, Academic Press Inc., San Diego, CA, 2001. With contributions by A. Brandt, P. Oswald and K. Stüben.

DIVISION OF APPLIED MATHEMATICS, BROWN UNIVERSITY, 182 GEORGE ST, PROVIDENCE, RI 02912, USA.

COMPUTER SCIENCE AND MATHEMATICS DIVISION, OAK RIDGE NATIONAL LABORATORY, OAK RIDGE, TN 37831, USA.

E-mail address: `Mark.Ainsworth@Brown.edu`

DIVISION OF APPLIED MATHEMATICS, BROWN UNIVERSITY, 182 GEORGE ST, PROVIDENCE, RI 02912, USA.

E-mail address: `Christian.Glusa@Brown.edu`